PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C12Q 1/68, G01N 33/53 C12N 5/00, 7/00, C12P 19/34

A1

(11) International Publication Number: WO 90/12891

(43) International Publication Date: 1 November 1990 (01.11.90)

(21) International Application Number: PCT/US89/01983

(22) International Filing Date: 9 May 1989 (09.05.89)

(30) Priority data:
337,990 14 April 1989 (14.04.89) US

(71) Applicant: GENMAP, INC. [US/US]; 291 Witney Avenue, New Haven, CT 06511 (US).

(72) Inventor: LANE, Michael, J. ; 6456 Tilden Drive, Syracuse, NY 13057 (US).

(74) Agents: MISROCK, S., Leslie et al.; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: METHOD OF PHYSICALLY MAPPING GENETIC MATERIAL

(57) Abstract

A DNA cassette is disclosed, which DNA cassette comprises a rare restriction sequence flanked by a unique DNA A sequence and/or a unique DNA B sequence. The DNA cassette can be inserted into genomic DNA. A method is also disclosed which uses said cassette to physically map the genomic DNA.

+ See back of page

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MC	Monaco
AU	Australia	FI	Finland	MG	Madagascar
BB	Barbados	FR	France	ML	Mali
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GR	Greece	NL	Netherlands
BJ	Benin	HU	Hungary	NO	Norway
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	SD	Sudan
CF	Central African Republic	KP	Democratic People's Republic of Korea	SE	Sweden
CG	Congo			SN	Senegal
CH	Switzerland	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
DE	Germany, Federal Republic of	LK	Sri Lanka	TG	Togo
DK	Denmark	LU	Luxembourg	US	United States of America

METHOD OF PHYSICALLY MAPPING GENETIC MATERIAL

TABLE OF CONTENTS

	Page
1. Field of the Invention	3
2. Background of the Invention	3
2.1. Pulsed Electrophoresis and Rare DNA Cutting Schemes	3
2.2. Size of Genomes as a Technical Analysis Problem	3
3. Summary of the Invention	6
4. Summary of Drawings	8
5. Detailed Description of the Invention	10
5.1. Cassette Construction	10
5.1.1. Rare Restriction Site	11
5.1.2. Unique DNA and Vector Selection	15
5.1.3. Cassette Integration	16
5.2. Propagation of Transformed Cells	17
5.3. DNA Fragment Preparation	18
5.3.1. Secondary DNA Treatment	18
5.3.2. Fragment Separation	20
5.4. Mapping Procedures	20
5.5. Mapping a Mammalian (Human) Genome	22
5.6. Constructing a Drosophila Map	23
5.7. Gene Localization	25
6. Constructing a Human Map	26
6.1. Vector Construction	26
6.2. Cassette Insertion	27
6.3. DNA Fragment Preparation	27
6.4. Pulsed Field Electrophoresis Separation	28
7. Deposit of Microorganisms	31

This application is a continuation-in-part of U.S. Patent Application Serial No. 06/315,017, filed on October 3, 1986, which is incorporated herein in its entirety by reference.

1. FIELD OF INVENTION

The present invention relates to novel DNA cassettes and a method of mapping genetic material in procaryotes and eucaryotes, including humans, using such cassettes. This method incorporates several technologies, including incorporation of synthetic and/or natural DNA sequences into genomic DNA, generation of rare restriction enzyme cutting sites, and size resolution of DNA fragments up to and greater than the million base pair size range. These technologies can be used to create a map of genomic DNA. Once a genomic map has been created, it can be compared to subsequently generated maps of cellular or organismal DNA. The method is useful in locating genetic lesions or alterations in the primary DNA sequence by comparison of such maps. This comparative method is thus capable of detecting genetic disorders, diseases, polymorphic loci (polymorphic alleles), and genetic alterations.

2. BACKGROUND OF THE INVENTION

2.1. PULSED ELECTROPHORESIS AND RARE DNA CUTTING SCHEMES

Over the past several years, the ability to resolve DNAs of large size has made it possible to consider restriction mapping and/or sequencing of the entire genome of an organism. The keystone to this advance has been the development of a related series of techniques which electrophorese DNA through gel matrices such as, but not limited to, agarose gels by employing pulsing electric fields (Schwartz, B. and Cantor, C., Cell 37:67, 1984; Snell and Wilkins, 1986). These techniques make it possible to resolve and analyze DNA fragment sizes orders of magnitude greater in size than was possible through historical electrophoretic techniques. The independent development of techniques to generate extremely rare restriction enzyme

cleavage specifities *in vitro* allows for the generation of large restriction fragments (McClelland *et al.*, *Proc. Natl. Acad. Sci. USA* 81:983-87, 1984; New England Biolabs Catalog 1985-1986, p. 29ff). The mechanism described therein relies on the use of a methylase which methylates DNA at specific sites recognizable subsequently by a restriction enzyme which cleaves DNA only when it is methylated at the restriction enzyme cleavage sequence. Appropriate choice of the correct methylation system therefore allows generation of very large restriction fragments.

2.2. SIZE OF GENOMES AS A TECHNICAL ANALYSIS PROBLEM

The problems encountered in analyzing large genomes cannot be overstated. For example, the human genome is approximately $3x10^9$ base pairs in length covering an estimated 3300 centimorgans (White *et al.*, *Nature* 313:101-105, 1985). Estimates of the number of marker loci needed to span this genetic length range upwards from 100 (Lange, K. and Boenhke, M., *Am. J. Hum. Genet.* 34:842-45, 1985; Botstein *et al.*, *Am. J. Hum. Genet.* 32:314-331, 1980). In a paper on how to generate human linkage, it has been stated, referring to the effort needed, "This will be a large scale endeavour and high efficiency of data collection will be important". (White *et al.*, *supra* at 101). Current estimates of how long it will take to construct a linkage map of human DNA range from 2 to 5 years assuming a great deal of effort from many researchers is combined (Lewin, R., *Science* 233:157-58, 1986).

An effective map suitable for general diagnostic and prognostic testing would require far more information than the 100 markers cited above. Ideally the map would have markers spaced every 50 kilobases of DNA or less, or would consist of upwards of 10^4-10^5 markers. Generation of this many ordered markers is not feasible using current techniques.

While advances have been made in constructing genetic maps by the implementation of various molecular biology techniques, at present count, less than a thousand genes, spanning only a small portion of the human genome, have been cloned (Willard et al., Cytogenet. Cell Genetics 80, 1985). These cloned genes have been used as probes to identify restriction fragment length polymorphisms (RFLPS) in genomic DNA and have proven useful in diagnosing some genetic disorders. Despite their utility, the information that can be gained through the use of such probes is limited by the fact that they span, at most, 1% of the human genome. In addition, it is necessary to isolate a different probe(s) for each disease. A clearly beneficial diagnostic tool would therefore be the ability to rapidly screen the entire human genome (Botstein et al., Am. J. Human Genetics 32, 1980) without the necessity of isolating probe(s) for each disease.

Accordingly it is an object of the instant invention to be able to screen an entire genome of a cell or organism rapidly.

It is an additional object of the instant invention to be able to generate a map or partial maps of a cell's or organism's genomic DNA.

It is another object of this invention to be able to comparatively assess the differences between genomic maps generated by this method.

It is a further object of this invention to use the comparative information thus generated to locate, identify, or define genetic lesions, mutations, insertions, deletions and other defects and polymorphisms in genomic DNA.

It is still a further object of this invention to use the instant invention as a diagnostic or prognostic tool in order to detect genetic defects, such as, but not limited to, prenatal diagnosis of genetic aberration, inherited genetic disease, and irduced or acquired genetic disease.

Still other objects and advantages will become readily apparent form the following description and claims.

3. SUMMARY OF THE INVENTION

The present invention involves a method and biological tools for mapping genomic DNA. This mapping technique can be used as a diagnostic test for detecting genetic disease and polymorphic loci and as a prognostic test.

The present mapping method comprises integrating DNA of the organism whose genome is to be mapped with a cassette of DNA containing a rare restriction sequence or site, the rare sequence or site being flanked on one or both sides by a unique DNA sequence. By "rare restriction sequence or site" it is meant one which does not occur, or occurs at very low frequency in the genome to be mapped, or can be made to be cleaved preferentially over genomic sites by any means rare, so long as its frequency allows partial mapping away from the rare restriction sequence into the genome of the host cell. The "unique DNA sequences", hereinafter referred to as DNA A, and optionally, DNA B, need not be restriction sequences, but rather are simply sequences capable of being identified uniquely in a background of host cell genome, which preferably do not occur in the host cell genome, and which differ from each other as well as from the rare sequence. In one embodiment, the cassette is inserted into the host cells by way of a vector, preferably a vector which will accomplish gene transfer through a single random integration of the cassette into the host cell. Independently derived,

monotonically integrated clonal isolates are then examined and analyzed. The unique DNA sequence(s) flanking the rare restriction sequence serves as a marker within the genomic DNA. The clonal cells are propagated independently, and then the restriction fragment pattern produced by each is examined. Fragments are generated by cutting at the rare restriction site, and at least one secondary restriction site, and then separated. The generated fragments are identified by virtue of the presence of the uniquely identifiable DNA sequences adjacent to the rare restriction site. Distance between the unique sequence and secondary sites on each fragment are measured, and then, by comparison, distances between restriction sites calculated. From this information, a regional restriction map can be constructed; repetition of this process ultimately permits construction of a total genomic map, by recognition of secondary restriction pattern overlap.

The invention also provides novel DNA cassettes comprising a restriction sequence rare for the genome to be tested, flanked on at least one side by a nucleotide sequence which is uniquely identifiable in the genome to be tested; also provided are novel vectors containing the cassettes, for use in integrating the cassette into the host cell genome. The nature of the cassette and vector will generally differ depending on the source of the genome to be mapped. In one embodiment, for example, any genome which does not have A methylation, and has an appropriate genomic utilization can be mapped utilizing a cassette comprising a rare ClaI/ClaI overalapping restriction sequence flanked on one or both sides by a retroviral sequence. Examples of organisms which would fall into such a category include but are not limited to mammals (humans), birds, and Drosophila; this cassette can be transmitted to the host cell by way of

a retroviral vector. Similar other constructs also can be created for mapping any other genomes, for example, other vertebrates, and invertebrates, yeast, plants, and bacteria.

The present invention also provides cell cultures or organisms, into the genome of which have been integrated the novel cassettes described above.

The method of the present invention provides many significant advantages over currently known mapping techniques. The present methods of choice, as noted above, rely on the identification of naturally occurring RFLP markers, and constructing linkage maps by pedigree analysis; this requires observation of numerous generations of individuals. Moreover, the average resolvable distance between RFLP markers at present is no better than several million base pairs. In contrast, the present method does not require an extensive pedigree study; also, resolvable distances are not limited by RFLP markers, but rather are dependent only upon the available cleavage and resolution technology.

The method provides high mapping accuracy with a rapidity heretofore uncontemplated in the art.

4. <u>SUMMARY OF DRAWINGS</u>

Figure 1. Flow diagram illustrating a genomic insertion mapping procedure for mapping mammalian cells. Each arrow indicates a step in the procedure with the expected DNA structures shown in boxed insets. The defective retroviruses and the <u>trans</u> packaging cell lines are described in Watanabe, S. and Temin, H.J., <u>Mol. Cell Biol.</u> <u>3</u>:2244-49, 1983; Mann <u>et al.</u>, <u>Cell</u> <u>33</u>:153-59, 1984; Sorge, <u>et al.</u>, <u>Mol. Cell Biol.</u> <u>4</u>:1730-37, 1984. The unique DNAs, A and B, correspond to the retroviral sequences at the 5' and 3' sides, respectively, of the Cla I/Cla I inserted sequence. This sequence, ATCGATCGAT, will be inserted in a

non-essential region of the defective retrovirus so as to allow subsequent replication and packaging by factors provided in trans.

Figure 2 shows maps of the vector pZipNeo indicating the sequence and position of the inserted oligonucleotide restriction site. Fig. 2a specifically shows the pZipNeo vector containing the ClaI/ClaI/DpnI sites; Fig. 2b illustrates a pZipNeo vector having multiple copies of the NotI recognition sequence.

Figure 3 schematically illustrates a procedure for locating the position of a particular gene or DNA segment once a map is constructed. Description of the method is found in Section 5.7, infra.

Figure 4 illustrates the presence and in situ MClaI/DpnI cleavability of insertion mapping vector pZipNeo28. Lane 1 - lambda Hind III markers; Lane 2 - MClaI/DpnI:SstI digest; Lane 3 - BamHI:SstI digest; Lane 4 - MClaI:ClaI digest control; Lane 5 - minus enzyme control; Lane 6 - ΨAM Neo minus control. The lower 3.3 kb band represents the actual neo gene while the two larger bands derive from the pBR322 section of transfected DNA. Note that both BamHI:SsT I digestion and MClaI:DpnI:SstI digestion produce the same band pattern indicating successful cleavage by the rare cutting strategy.

Figure 5 illustrates the use of lambda phage concatemers as pulsed field electrophoresis size markers with whole yeast chromosomes on the outside lane as a reference. Twelve distinct bands are resolved in the outside lanes in the figure containing all 17 chromosomes (6 bands represent doublets).

5. DETAILED DESCRIPTION OF THE INVENTION

5.1. CASSETTE CONSTRUCTION

In accordance with the present invention, genomic DNA can be mapped by inserting into said genomic DNA a DNA sequence which is a rare cleavage site in the context of the host DNA with which it is integrated. This rare restriction sequence is flanked at one end by a uniquely identifiable DNA sequence, termed unique DNA A, and optionally at the other end by a second uniquely identifiable DNA sequence, termed unique DNA B. These unique DNA sequences can be natural or synthetic. This combination of sequences, i.e., the rare restriction sequence, unique DNA A and optional unique DNA B, is termed a cassette. Optionally the cassette may contain a sequence or sequences which facilitate subsequent isolation of the genomic DNA flanking the cassette. For example, the cassette may also contain a high affinity protein binding site. In one embodiment, the λ repressor binding sequence can be used in conjunction with a DNA affinity column composed of covalently bound repressor monomers. In this way, DNA fragments containing the unique DNA A or unique DNA B sequence can be readily isolated for subsequent manipulation. In another embodiment the cassette can contain genetic functionalities that allow it to be maintained as a plasmid in E. coli or other appropriate host, thus facilitating ready isolation of the unique DNA A or unique DNA B and flanking genomic DNA for subsequent manipulation.

The actual sequences of the rare restriction sequence, unique DNA A, and unique DNA B are not critical. These sequences, however, should be different from one another, and, in a preferred embodiment, the sequences should occur infrequently, if at all (i.e., they are underepresented) in the host genomic DNA. It is possible that a similar sequence or sequences exists in the host

organism. All that is required in this procedure is to differentiate between the inserted DNA and the endogenous DNA.

5.1.1. RARE RESTRICTION SITE

The identity of the rare restriction site will differ depending upon the host organism whose genome is to be mapped, since a particular sequence may be rare in one organism, but not another. The term "rare" in the present context can best be defined operationally. An initial consideration in choosing an appropriate sequence is what will be the preferred fragment size resulting from cleavage. The preferred fragment size is not dictated by any particular requirement, but rather is a matter of convenience: the larger the fragment size produced, generally, the easier the mapping procedure will be. "Large" is, of course, determined relative to the total size of the genome to be mapped. Smaller fragments are just as acceptable functionally, but require many more repetitions of the procedure in order to get an equivalent map. For this reason, large fragments are preferred.

Once a general determination is made as to a fragment size which would be acceptable for the purposes of the genome under consideration, the choice of a rare cutter can be made in a number of ways. One approach is to simply treat the DNA with an appropriate enzyme (appropriate to be defined below), and observe the size of the fragments produced. If the fragment size is acceptable in accordance with the guidelines noted above, then a useful sequence has been chosen, and may be used in the present procedure.

On the other hand, a more systematic approach to the selection of a sequence may be desired. In such a case, an appropriate sequence may be predicted empirically by reference to the overall nucleotide composition of the genome to be mapped. For example, a general knowledge of

the approximate GC content of the genome provides a convenient means by which the expected average fragment size, in base pairs, generated by cleavage at any given restriction site can be predicted. Information relating to GC content of various organisms is readily available in the literature (e.g., Hill, J. Gen. Microbiol. 44:419-437, 1966), or is readily determinable by known techniques (Owen, R.J. and Pitcher, D. (1985) In "Chemical Methods in Bacterial Systematics" pp. 1-15, Edited by M. Goodfellow and D.E. Minnikin, Academic Press, London). Given the fraction of total DNA which is GC, AT content can also be determined (fraction GC + fraction AT = 1). Assuming random order of dinucleotides/trinucleotides, then average fragment size (AFS) generable by cleavage of a given recognition sequence can be calculated by the following formula

$$AFS = \left(\frac{2}{r_1}\right)^a \left(\frac{2}{1-r_1}\right)^b$$

where r_1 = fractional GC content
$1-r_1$ = fractional AT content
a = # G + C in recognition sequence
b = # A + T in recognition sequence.

For example, assume .40 G+C content and .60 A+T content. The sequence of choice is ATCGAT. In this case:

$r_1 = .4$ $a = 2$
$1-r_1 = .6$ $b = 4$

Inserting these values in the formula,

$$\left(\frac{2}{.4}\right)^2 \left(\frac{2}{.6}\right)^4 = (25)\ (123.5) = 3086 \text{ base pairs}$$

Thus, the average fragment size produced by cleavage of the restriction sequence ATCGAT in this genomic environment is estimated to be 3086 base pairs. Given the initial estimation of the desired fragment size for the genome of choice, it is readily apparent whether or not the chosen site is acceptable for the purpose.

The above schemes are not the only methods by which an appropriate restriction sequence can be chosen, but modifications thereof will be readily apparent to those skilled in the art. Similar equations have been previously described (e.g., Nei and Li, PNAS USA 76:5269, 1979). Also, summaries of rare v. common sequences are available in the literature (McClelland et al., in Gene Amplification and Analysis, Chirikjian (ed.), Elsevier Science Publishing Co., 1987, pp. 258-282, and references cited therein). Thus, the skilled artisan can routinely make an appropriate selection of a rare restriction site for the genome in which he is interested.

In one embodiment, the selection of a rare site can be taken an additional step, by modification of the sequence in a manner which renders it even less likely to be cut than it would be in its unmodified state. Selective methylation of a particular sequence, for example, may, depending on the organism, result in the production of a highly specific cleavage site which is only rarely cut in the genome of choice (McClelland et al., PNAS USA 81:983-987, 1984).

Alternately, the chosen cleavage site can be arranged in tandem arrays. This will normally result in a preferential cleavage of the chosen site within the cassette, over cleavage of the same sequence in a genomic

site, in turn producing greater efficiency in large fragment production. Similarly, incorporation of a DNA molecule modified to cleave, such as a D-loop, or a triple helix (Strobel et al., J. Am. Chem. Soc. 110:7927-7929, 1988) will achieve substantially the same effect. These are but a few examples however, and any modification of the chosen sequence which ultimately aids in the generation of appropriate fragment sizes by selectively concentrating cleavage at the site of cassette insertion is contemplated.

In the case of modification of the sequence selected, the initial sequence need not even be particularly uncommon in the host genome, but merely need to be modifiable in such a way as to render them "rare" in the present context.

In one embodiment, when a human genome is being mapped, a preferred sequence can be the overlapping ClaI/ClaI sequence

ATCGATCGAT
TAGCTAGCTA.

This site is of particular utility since it is subject to selective methylation by the enzyme MClaI (McClelland et al., PNAS USA 81:983-987, 1984). This methylation renders a rare 10-base sequence cleavable, since mammalian DNA is not routinely methylated at ClaI. The methylated 10-base ClaI sequence is subject to selective cleavage by the restriction endonuclease DpnI (or CfuI), which cuts only DNA which is methylated at adenine in both strands of the recognition site. An additional benefit can be obtained by constructing this rare site, within the cassette, in tandem repeats. Surprisingly, the efficiency of cleavage within the cassette appears more than additively increased when compared with cassettes containing a single copy of the sequence.

The selected oligonucleotide restriction sites can be readily prepared synthetically.

5.1.2. UNIQUE DNA AND VECTOR SELECTION

The purpose of the uniquely identifiable DNA flanking the rare restriction site is to provide a basis for detecting the cassette amidst the genomic DNA. In order to fulfill this purpose, unique DNA A and unique DNA B need only be distinguishable, by some detectable means, from the host DNA. To this end, one may synthetically generate sequences which are, based on knowledge of the overall composition of the host genome, expected to occur only rarely, if at all, in the genomic DNA. Alternately, the sequences can be generated by fragmentation and isolation of genomic DNA derived from an organism genetically distant from the host organism. For example, for mapping eukaryotic genomes, unique DNA can be derived from procaryotic i.e., bacterial or viral, genomic DNA. The unique DNA is then detectable by virtue of, for example, hybridization with a labelled complementary DNA probe, or the presence of a selectable marker.

In a preferred embodiment, the unique DNA sequences are chosen in association with a vector used to transform the host cells. In other words, the vector chosen is preferably one sufficiently distinct genetically from the host cell to permit detection of the vector DNA after its integration into the host cell genome. For example, in one embodiment a replication-defective amphotrophic virus vector, into which the rare restriction sequence has been inserted, such as that described by Sorge et al. (Mol. Cell Biol. 4:1730-1737, 1984), is used to infect a mammalian cell line. These viruses are capable of infecting cells, but once genomically integrated, are incapable of post-insertion replication, preventing reinsertion by the virus into other segments of the genome.

5.1.3. CASSETTE INTEGRATION

The cassette constructed, as outlined above, must be integrated with the genomic DNA to be mapped. This integration can be achieved by any method useful in attaining DNA transfer; this includes, but is not limited to, the use of electroporation, micrinjection, infection or ligation into a cloning vector. The term "integration" in the present context means the association of the cassette with the genomic DNA in a continuous piece of DNA.

In a preferred embodiment, however, the cassette is integrated into the genomic DNA to be analyzed by use of a vector which inserts the cassette into a host cell. For the present purposes, the vector is preferably a transposon-like element, i.e., one capable of being integrated into the genome essentially "at will". The use of a vector provides a convenient source of uniquely identifiable DNA: insertion of the rare restriction sequence into a properly chosen vector automatically flanks the restriction sequence with a distinct sequence which, upon integration into the host cell genome, will be readily detectable, provided the vector sequence is sufficiently distinguished from the host cell.

Appropriate vectors for a variety of different cell types are readily available. For example, for DNA insertion into prokaryotic cells, the utility of transposable elements has long been recognized (Kleckner, Cell 11:11-23, 1977; Calos et al., Cell 20:579-595, 1980), and provide a means for sequence neutral integration of the type necessary to attain insertion of the restriction site. Ty elements of yeast are also similar in structure, and to some extent, function, to the prokaryote transposons (Boeke et al., Cell 40:491-500, 1985). In Drosophila, P elements have been routinely used to introduce cloned sequences into the organism (Steller et al., EMBO J. 4:167-171, 1985). In higher plants, the use of Ti plasmids to introduce exogenous

DNA is now relatively routine technology (Chilton, M. et al., Cell 11:263-271, 1977) and plant virus vectors may also be used. For integration of DNA into mammalian cells, viral vectors can be employed. Particularly preferred as a vector for mammalian cells are retroviral vectors. Alternate choices for vectors will be readily apparent to those skilled in the art.

The trait of being "uniquely identifiable" is intended to convey that, by some means, the presence of the cassette by detection of the unique DNA, can be verified. A convenient method or achieving this is by the use of a vector which is genetically distant from the host; in this way, insertion of the cassette can be verified by hybridization with a vector-specific probe, which probe will not hybridize with the nonhomologous host genomic DNA. Alternately, the cassette can be constructed so as to include a particular selectable marker which allows the identification of the presence of cassette DNA.

5.2. PROPAGATION OF TRANSFORMED CELLS

In the case in which the cassette DNA is integrated into a host cell, following integration into the genomic DNA, the host cells are divided into single cell solutions. This can be accomplished by, for example, dilution or cell sorting. The cells are then propagated in a manner consistent with culture conditions required for the the cell line selected. Organisms will be treated by whatever means necessary so as to effect the same result.

For example, when using a viral vector to integrate cassette DNA into mammalian cells, the cells are serially diluted in order to produce cultures containing a single cell. These clones are then propagated in the presence of a selectable agent which will prevent growth of cells which have not integrated a copy of the virus. (See, e.g., Watanabe, S. and Temin, H.J., Mol. Cell Biol. e:2244-

2249, 1983; Mann et al., Cell 33:153-159, 1984; Sorge et al., Mol. Cell Biol. 4:1730-1737, 1984). This can be achieved by insertion, with the vector, of a selectable marker, such as antibiotic resistance. Similar screening procedures can be achieved with whole organisms, e.g., whole plants. Here, vectors can also be constructed to carry a selectable marker, and plant cell cultures transformed thereby. Plants regenerated from culture can be screened on a selective medium at an early stage of development, and the surviving plants represent those which have integrated cassette DNA.

The above technique will result in clonal populations which have an extremely high probability of containing a single inserted cassette. A single insertion is preferable, but not critical, to the present method. Insertion of additional cassettes, of different structure from the first, is also contemplated. Optionally, one can, at this point, verify that each clonal population does contain a single insert of a given construction via a variety of conventional techniques, such as assaying relative cassette copy number per cell by comparison with a known cassette DNA concentration via hybridization. It is understood, however, that any technique which will identify the clonal lines which contain a single insertion of the cassette is acceptable for the instant invention. It is also understood that insertion of the cassette into the genomic DNA, causing integration of the cassette and genome, is only one means of attaining integration.

5.3. DNA ISOLATION

The clonal populations which contain the genomic DNA to be mapped are then lysed in any manner which is suitable for the DNA separation method selected. These techniques can include, but are not limited to, prior suspension of cells in agarose, e.g., agarose microbead

technique (Cook, P. EMBO Jour. 3:1837, 1984) and agarose block technique (Schwartz, B. and Cantor, C., Cell 37:67, 1984 and U.S. Patent No. 4,473,452). These techniques allow in situ cell lysis by enzymes, detergents and proteins diffused into the agarose while maintaining DNA integrity. Any method of DNA isolation which leaves the DNA in a state available for subsequent treatment is acceptable (see, e.g. Maniatis et al., supra).

5.3.1. SECONDARY DNA TREATMENT

After isolation of DNA, the genomic DNA is then treated so as to produce fragments suitable for mapping. In general, the DNA will be cleaved with a restriction enzyme having specificity for the rare restriction site, and at least one secondary restriction enzyme. In one embodiment, as already noted above, the rare restriction site is first treated with a site-specific methylating enzyme, in order to render the restriction site more rare, and then followed by cutting with methylation dependent restriction enzyme. In accordance with a overall strategy of the present method, in a preferred embodiment this initial enzyme treatment will preferentially cut within the cassette, and will produce little or no cleavage within the genomic DNA.

The DNA is also partially digested with one, or independently, a series or mixture of secondary restriction enzymes, so that each DNA sample will be digested with the restriction enzyme specific for the rare site, and partially digested with the secondary restriction enzymes. These secondary enzymes will be specific for various sequences within the genomic DNA. The identity of the enzyme used is not critical, and can be any restriction enzyme which cuts within the genome to be mapped. However, if it is desired to produce larger fragments for mapping, the chosen enzyme will preferably be one which cuts a relatively uncommon

site. Many such enzymes are available as commercial products, such as, for example, XmaIII and XmnI (New England Biolabs, Inc.)

5.3.2. FRAGMENT SEPARATION

Following these enzymatic reactions, the genomic DNA restriction fragments are separated, either according to size or molecular weight. This separation may be achieved by any method which is capable of resolving fragments of the size produced. A majority of the current techniques rely on electrophoretic separation. The choice of technique will to a large extent govern the ultimate resolution that can be obtained in the mapping procedure. Any technique that allows measurement of the distance from the rare site to the mapping site is acceptable. For example, HPLC methods are not particularly suited to separation of large fragments. In a preferred embodiment of the present method, fragments are separated via pulsed field electrophoresis, as described, for example in Schwartz and Cantor, Cell 37:67, 1984, and U.S. Patent No. 4,473,452. This technique is particularly well suited to separation of the large fragments which are preferred for this method. In this technique, the genomic fragments are contained in a suitable medium, preferably a gel medium and the DNA fragments subjected to pulsing electric fields.

5.4. MAPPING PROCEDURES

Once the genomic fragments have been separated, those fragments which contain the unique DNA A/DNA B sequences are identified. This can be achieved by any method capable of distinguishing the unique cassette DNA from the genomic DNA background of the host cell organism. A convenient method for specifically identifying the unique sequences is by probing the fragments with a unique cassette DNA-specific, labelled, cloned DNA fragment substantially

homologous to one of the unique flanking DNA sites. Preferably the complementary sequence will be labelled with a radioactive, fluorescent or color indictor. Most frequently, the separated DNA fragments will be blotted onto a support membrane, such as, but not limited to, nylon or nitrocellulose, prior to hybridization, in accordance with the method of Southern. This procedure results in an end-label of only those fragments containing the introduced restriction site, producing a ladder of fragments giving the genomic restriction site pattern away from the integration site in one direction. The same blot can be probed with a sequence homologous to the other side of the restriction site, (i.e., DNA B, if present), producing a fragment pattern representing the genomic restriction sites on the other side of the integrated cassette.

Once the restriction fragments have been generated and physically ordered, physical mapping can be initiated. The general strategy employed for physical mapping is a contig strategy which has been described previously for mapping of the yeast and nematode genomes (Olson et al., PNAS USA 83:7826, 1986; Carlson et al., PNAS USA 83:7821, 1986; Carlson et al., Nature 335:184, 1988). In very general terms, the mapping procedure is as follows:

One fragment will represent that portion of the genomic DNA from unique DNA A to a first secondary restriction enzyme cut. A second fragment will represent the distance from unique DNA A to a second secondary restriction enzyme cut. Deducting the first distance from the second distance generates the distance from the first to the second secondary cuts. For large genomes, the above procedure will have to be repeated a number of times, the number of times being dependent on the length of the genome to mapped. The clonal maps are then compared in order to

ascertain overlapping portions. By aligning these overlapping portions, a complete map of the genomic DNA can be generated.

5.5. CONSTRUCTING A MAMMALIAN (HUMAN) MAP

A brief summary of the procedures followed in this method is found in Figure 1. Briefly, the ClaI/ClaI overlapping sequence ATCGATCGAT has an estimated frequency of occurrence in the human genome of once every 2 x 10^8 base-pairs. This sequence is inserted into the DNA of a defective amphotropic retrovirus. The DNA recombinant is then transfected into a cell line harboring a trans acting, replication defective copy of the retrovirus (See, e.g., Sorge et al., Mol. Cell. Biol. 4:1730, 1984; Cohn et al., PNAS 63:49, 1981). This allows assembly of RNA containing viral particles, which are then exported from the cell. This type of construction has the demonstrated capability to infect cell lines, but is incapable of post-insertional replication. These particles are used to infect, monotonically, untransformed cells which contain the genome of interest. Infected cells, containing a copy of the virus, are clonally propagated. As proviral insertion is an essentially random process, each of the individual clones will have its genome uniquely marked by a defective virus containing the ClaI/ClaI sequence. To construct a physical map of the DNA surrounding the retroviral integration site in a given clone, DNA is isolated and then methylated with M.Cla I (McClelland, Nucl. Acids Res. 9:6795-6804, 1981), creating a DpnI site at the overlapping ClaI/ClaI site. Following cleavage with DpnI, the DNA is partially digested with a second restriction enzyme and then electrophoresed next to appropriate DNA size markers, for example, pulse field electrophoresis with a partially annealed λ phage ladder as size standard. In this approach, electrophoresis is carried out under conditions which allow resolution of

large DNA molecules in order to acquire as much mapping information as possible. If the gel is then southern blotted and probed with nick translated, cloned, viral DNA from one side of the *ClaI*/*ClaI* introduced site, this will result in an end label of only those fragments containing the introduced *ClaI*/*ClaI* site, producing a ladder of fragments giving the restriction pattern away from the viral integration site in one direction. Subsequently, a probe from the other side of the *ClaI*/*ClaI* site is used to produce a ladder of fragments representing the restriction pattern of the genomic DNA on the other side of the integrated provirus. In this fashion, at least 3-4 million base-pairs may be restriction mapped for each clonal cell isolate. Assuming the human genome is 3 billion base-pairs in length (haploid) it would take approximately 750 independent isolates to cover the genome to a map density of 1, wherein map density is defined as the number of genome equivalents in DNA base pairs mapped. If 50 lanes may be run on each gel, such a map could be created by running 30 pulsed field gels. Once the first map is created, mapping can be done comparatively.

5.6. CONSTRUCTING A DROSOPHILA MAP

A *BamHI* sticky ended *ClaI*/*ClaI* oligonucleotide,

5'-GATCCATCGATCGATG-3'
3'-GTAGCTAGCTACCTAG-5',

was synthesized and inserted in tandem arrays into the retroviral vector pZipNeo as described above for mammalian mapping. Using the same strategy the same oligonucleotide can be inserted into the P element vector, pUChsneo (Steller and Pirotta, *EMBO J* *4*:167, 1985). The polylinker cloning site of pUChsneo is opened with *BamH1*, the phosphorylated linker added, and closed with ligase. The ligation products are used to transform the *E. coli* strain GM2929 (dam -).

Selection for insert positive transformants will follow the strategy of Viera and Messing (Gene 19:259, 1982). Positive clones are grown up and insert size (e.g., number of ClaI/ClaI sites inserted) determined by double digestion of each positive plasmid with SmaI and SalI. Each has unique sites flanking the insert (Steller, H. and Pirrotta, V., supra). The insert sizes are then resolved on a native polyacrylamide gel against size standards. Plasmids containing 6, 8, and 10 ClaI/ClaI sites, respectively are grown up preparatively by standard techniques (Maniatis, Molecular Cloning, Cold Spring Harbor Laboratory, 1982). These plasmids are used to microinject Canton S embryos.

Embryos obtained from a Canton S (Brown University) strain are injected (Zalokar, Microscopica Acta 84:231, 1981; Zalokar, Experientia 37:1354, 1981; Santamaria, Dev. Bio. 96:285, 1983) with the P element construct generated as described above along with the transposase positive pπ25.7wc (Karess and Rubin, Cell 38:135, 1984). Positive P element containing flies are selected by addition of G418 to the growth media as described (Steller and Pirotta, supra). Positive fly stocks are maintained under G418 selection for several generations in order to ensure stable P element insertion into the germ line. P element insertion can be verified by isolation of DNA from each positive line and spot blotting (Kafatos et al., Nucl. Acid Res. 7:1541, 1982) followed by hybridization with [32]P labeled P element plasmid.

Each isolated transformed line is maintained for production of DNA for restriction mapping from the inserted P element. DNA can be prepared from overnight collections of embryos which have been dechorionated (Santamaria, in Drosophila, A Practical Approach, D.B. Roberts, ed., IRL Press, 1986) prior to embedding in agarose plugs (Schwartz and Cantor, Cell 37:67, 1984) for lysis. Lysis is performed in a Tris buffer containing 1 mg/ml proteinase K and 1%

sarcosine (5-10 µl) of each agarose plug containing the now-purified DNA is then rinsed in buffer composed of 10 mM Tris/1 mM Na2EDTA for one hour, then the agarose slice washed in 10 ml of the same buffer containing 200 µl of 100 mM PMSF (phenylmethyl sulfonyl fluoride). This procedure is repeated once again, then the PMSF is washed out of the slice by a one-hour incubation in the 10 mM Tris/1 mM EDTA solution. The DNA is then clean enough for further manipulation.

5.7. GENE LOCALIZATION

The method also has utility in identifying the position of any genomic DNA fragment, for example, to locate the map position of a cloned DNA segment. From rudimentary localization of contigs determined by in situ hybridization, it is possible to determine specific localization of a gene or genes known to fall in given chromosomal regions by cleaving the cell lines representing the appropriate contig to completion with both MClaI/DpnI and NotI. The procedure is outlined in Figure 3. DNA so cleaved can be pulse-electrophoresed, blotted and probed sequentially with probes hybridizable to the unique DNA to either side of the MClaI/DpnI site. This identifies MClaI/DpnI-NotI fragment sizes. The same blot can then be probed with the gene of interest. This procedure will identify the cell line in which the genomic NotI-NotI fragment containing the gene has been interrupted by retroviral insertion, and to which side of the MClaI/DpnI site the gene falls. This localizes the gene to that region of the contig. Failure to identify such a sequence would suggest repeating the procedure substituting a lower frequency cutting system for NotI. (e.g., MTaq/DpnI; McCleland and Nelson, supra).

The following example illustrates one method for generating maps of human DNA. It is understood, however, that the instant invention is not limited to human cells

only. Rather, genomes from any prokaryotic or eukaryotic cell or organism can be mapped using the instant invention, and the adaptations required will be readily recognized by those skilled in the art, in light of the specification and examples.

6. CONSTRUCTING A HUMAN MAP

6.1. VECTOR CONSTRUCTION

For use in mapping a human genome, the strategy is to utilize a defective retroviral vector to insert a rare restriction site into random positions in the human host cells. The rare restriction site selected is the ClaI/ClaI overlapping sequence, which has been estimated to occur at a frequency in the human genome of about once every 200,000,000 base pairs (McClelland and Nelson, *supra*). The following overlapping oligonucleotide is synthesized, flanked by BamHI cohesive ends:

5'-GATCCATCGATCGATG-3'

3'-GTAGCTAGCTACCTAG-5',

The oligonucleotide is inserted, both singly and in tandem arrays, at the unique BamHI into the murine (MuLV) retroviral shuttle vector pZipNeo originally described by Cepko (*Cell* 37:1043, 1984). This vector contains a pBR322 origin of replication, an SV40 origin of replication, and a selectable marker, the resistance gene for G418 (neomycin). Figure 2 shows the map of the vector pZipNeo. A number of constructs can be made: a vector containing a single copy of the oligonucleotide inserted (pZipNeo28; n=1); an identical vector with a tandem array of three oligonucleotides self-ligated and then inserted; (pZipNeo84; n=3); and a vector with six tandem oligonucleotides inserted (pZipNeo168; n=6). Tandem arrays of the sequence are created by ligating the insert to itself in the presence of T4 kinase, ^{32}P ATP and

T4 ligase. Products corresponding to 3 and 6 ligation events are isolated by elution from an 8% polyacrylamide gel following autoradiography.

6.2. CASSETTE INSERTION

The recombinant DNA's are then transfected into a cell line harboring a trans-acting, replication defective copy of the retrovirus (ψ AM) which allows assembly of recombinant RNA containing infective viral particles. pZipNeo DNA can be transfected by, for example, the scrape loading technique (Fechheimer, PNAS USA 84:8463, 1987) into the amphotrophic packaging line. The viral particles produced are used to infect, monotonically, clonal human embryo fibroblast line MRC-5, by incubation of the packaging cell line media with the fibroblast cells (Cohn & Mulligan, supra).

6.3. DNA FRAGMENT PREPARATION

Treated cells are selected for neomycin resistance by culturing with G418 for a period of 2-3 weeks. Clones are evident at this time, and are subsequently picked and grown up in the presence of G418. The cells are harvested at confluence and cast in 1% low melting agarose (FMC Corporation) and lysed to make "plugs" according to the method described in Schwartz and Cantor, supra. Plugs are treated for MClaI/DpnI digestion, first by cutting into 1/4 pieces. They are twice washed with Tris/EDTA buffer, then twice in Tris/EDTA buffer containing 1 mM PMSF and finally twice more in Tris/EDTA buffer. Plugs then are cut into 1/3 slivers and washed in microfuge tubes with Tris/EDTA containing 1 mM s-adenosylmethionine for 1 hour. MClaI reactions were incubated at 37°C, usually overnight, in the presence of 40% glycerol, 1 mM SAM, 5 mM DTT in Tris EDTA (pH 7.5). The next morning, all reaction buffer was replaced, additional M ClaI units were added and the

reaction continued for 4 hours. DNA slices are then washed briefly in several volumes of Tris/EDTA, then in DpnI buffer for 1/2 hour. DpnI reactions proceeded with 25 units of enzyme for 2-4 hours at 37°C. Reactions were halted with addition of ESP buffer and incubation at 55°C for 20 minutes.

Subsequent digestion reactions with secondary restriction enzymes are also performed after a washing protocol identical to that used prior to MClaI/DpnI (digestions without addition of s-adenosylmethionine to the wash buffers). Appropr'ate reactions would be then initiated and allowed to proceed overnight at the optimum temperature to the extent necessary to effect partial degradation of genomic DNA at the optimum temperature. Subsequently, the reactions are terminated, by addition of fresh lysis solution (e.g., 1 mg/ml Proteinase K/10 m<u>M</u> Tris (pH 9.0)/0.5 M disodium EDTA/1% sarcosine), and then pulse-electrophoresed on a 1% agarose gel, then blotted and probed with a XhoI-XhoI (Neo) probe created from pZipNeo DNA by random priming (BRL) of the fragment.

Results of a specific application of these procedures in which the secondary restriction enzyme was allowed to proceed to completion, the reactions terminated, and the samples loaded on a constant field electrophoresis device, electrophoresed and probed are shown in Figure 4.

6.4. <u>PULSED FIELD ELECTROPHORESIS SEPARATION</u>

In general, however, following these treatments the generated genomic DNA fragments will be separated by pulsed field electrophoresis as described in Cantor <u>et al.</u>, <u>supra</u>. Lambda phage concatemers are employed as electrophoresis size markers. Wild type whole lambda phage (Ci857sam7) are dialyzed after purification on a cesium chloride gradient and subsequently diluted in PBS and mixed with an equal volume of 1.5% of low gelling agarose (FMC lot

#12276). The molten agarose solution is mixed and then pipeted into plastic forms and allowed to solidify making agarose 'plugs' (Schwartz, D.C. et al., CSHSQB 47:189, 1982). Agarose plugs are suspended in a solution composed of 0.5 M disodium EDTA/1% sarcosine/1 mg/ml proteinase K/10 mM Tris-Cl (pH 9.0) and incubated at least 4 hours at 55°C with gentle shaking. Samples are loaded as described (Schwartz, D.C. et al., supra) onto a 1% agarose gel for electrophoresis. The gel shown was run for 62 hours at 8.5 V/cm with a pulse time of 150 seconds on an apparatus made in this laboratory after a modified design of Schwartz and Cantor. (Waterbury and Lane, Nucl. Acids Res. 15:1940, 1987). Final DNA concentrations of 0.06-0.60 µg/µl were utilized to illustrate the optimum concentration of phage DNA for good molecular weight "ladders". This technique is illustrated in Figure 5, using whole yeast chromosomes against a lambda phage ladder. Whole yeast chromosomes are prepared in plugs as described, with the exception of spheroplasting with zymolase at a concentration of 2 mg/ml prior to suspension in molten agarose solution. This figure shows a resolution of up to 1250 kb.

The separated, labelled blot of the fragments are exposed to film, the film used to assign molecular weights and the order in which they appear to the fragments, and then examined to recognize restriction pattern overlaps, from which a complete genome map can be determined.

The above invention will find many appplications and uses. For example, this invention provides for a fast and easy way to generate maps of complex genomes, including human genomes.

Maps of genomes can be compared to each other in order to detect any differences between them. For example, a genome can be mapped and compared to a standard map of that genome. This procedure will be important in such areas

as prenatal diagnosis of inherited genetic disease, identification of induced (acquired) genetic disease, as well as in prognostic applications.

In one embodiment, a genomic map is generated and placed in a data base. Thereafter, any other genomic maps generated are compared t the first map by use of an appropriate computer algorithm.

The map can also be used to locate specific genes, and to identify normal genes.

In another example, genomes from various cells in the same organism can be mapped and compared to detect for differences between them. This will allow for greater specificity, since most, if not all, of the genomes should be identical to each other, and detailed maps can be generated without having to account for variability. For example, a standard map can be prepared from a normal human cell, and that can be compared to map of a neoplastic cell from the same individual. This procedure will indicate what genetic changes a human cell undergoes as and when it becomes cancerous.

The technique can be used to create a library of marked cell lines, or marked whole organisms (plant or animal), each of which represent a particular part of the genome.

This invention will also find ready application in other fields, such as anthropology and evolutionary biology. Maps of genomes from various organisms can be generated and compared in order to further the study of evolution. Other fields will benefit as well, such as horticulture, animal husbandry and genetic engineering.

The present method also has particular significant advantage in its ability to effect map closure by non-random extension, at lower resolution, of maps produced from contig ends. The inability to close a map is

a drawback of other types of mapping techniques, and, in fact, the present method can be used to close maps prepared by these other methods.

7. DEPOSIT OF MICROORGANISMS

The pZipNeo vector containing a ClaI/ClaI overlapping restriction sequence is deposited, in an *E. coli* host, with the NRRL and has been assigned the accession number B-18490.

It is understood that the above uses for the instant invention are set forth as examples only. They are not meant to be limiting on the instant invention.

International Application No: PCT/ /

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page 31, line 5-13 of the description

A. IDENTIFICATION OF DEPOSIT

Further deposits are identified on an additional sheet [X]

Name of depositary institution

Agricultural Research Culture Collection

Address of depositary institution (including postal code and country)

1815 North University Street
Peoria, Illinois 61604

Date of deposit	Accession Number
April 28, 1989	B-18490

B. ADDITIONAL INDICATIONS (leave blank if not applicable). This information is continued on a separate attached sheet [ ]

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)

D. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. [X] This sheet was received with the international application when filed (to be checked by the receiving Office)

Virginia L. Lily
(Authorized Officer)

[ ] The date of receipt (from the applicant) by the International Bureau

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

Form 134 Continued

Name of Depository Institution: Agricultural Research Culture Collection

Address of Depository Institution: 1815 North University Street
Peoria, Illinois 61604

Date of Deposit: April 28, 1989

Accession Number: B-18489

WHAT IS CLAIMED IS:

1. A method of mapping genomic DNA which comprises

(a) integrating a DNA cassette with the genomic DNA, said DNA cassette comprising a rare restriction sequence flanked at one end by a uniquely identifiable DNA A sequence, and optionally at the other end by a uniquely identifiable DNA B sequence, wherein each sequence is different from the others, and wherein they are distinguishable from the genomic DNA;

(b) cutting the genomic DNA at the rare restriction sequence, and at least one secondary restriction sequence to produce fragments;

(c) identifying fragments containing a uniquely identifiable sequence;

(d) measuring distance between the uniquely identifiable sequence and the secondary restriction site;

(e) calculating therefrom distances between different secondary restriction sites; and

(f) generating a map therefrom, the map comprising the distances between secondary restriction sites.

2. The method of claim 1 wherein the fragments produced in step (b) are separated by pulsed field electrophoresis.

3. The method of claim 1 wherein the genome to be mapped is a vertebrate genome.

4. The method of claim 1 wherein the genome to be mapped is a mammalian genome.

5. The method of claim 1 wherein integration is achieved by insertion of the cassette into the genomic DNA of a cell.

6. The method of claim 3 wherein the genome to be mapped is a mammalian genome.

7. The method of claim 6 wherein the insertion is achieved by vector transmission of the cassette.

8. The method of claim 7 wherein the vector is a virus.

9. The method of claim 8 wherein the vector is a retrovirus.

10. The method of claim 9 wherein the retrovirus is an amphotropic retrovirus.

11. The method of claim 10 wherein the vector is a pZipNeo.

12. The method of any one of claims 3-11 wherein the rare sequence is a ClaI/ClaI restriction site.

13. The method of claim 12 wherein the ClaI site is treated with MClaI prior to cutting.

14. The method of claim 13 wherein the rare sequence is cut with DpnI.

15. The method of claim 1 wherein the rare sequence is present in a tandem array.

16. The method of claim 12 wherein the rare sequence is present in a tandem array.

17. The method of any one of claims 3-11 wherein the rare sequence is a NotI site in tandem array.

18. The method of claim 1 wherein the rare sequence is a D loop.

19. The method of claim 1 wherein the rare sequence is a triple helix.

20. The method of any one of claims 3-11 or 15 wherein the genomic DNA is human genomic DNA.

21. The method of claim 12 wherein the genomic DNA is human genomic DNA.

22. The method of claim 13 wherein the genomic DNA is human genomic DNA.

23. The method of claim 14 wherein the genomic DNA is human genomic DNA.

24. The method of claim 17 wherein the genomic DNA is human genomic DNA.

25. The method of claim 18 wherein the genomic DNA is human genomic DNA.

26. The method of claim 19 wherein the genomic DNA is human genomic DNA.

27. The method of claim 1 wherein the genome to be mapped is an insect genome.

28. The method of claim 27 wherein the DNA is integrated by insertion of the cassette into a cell containing the genomic DNA.

29. The method of claim 28 wherein insertion is achieved by vector transmission of the cassette.

30. The method of claim 29 wherein the vector is a P element.

31. The method of claim 30 wherein the vector is pUChsneo.

32. The method of claim 30 or 31 wherein the rare sequence is a ClaI/ClaI restriction site.

33. The method of claim 32 wherein the ClaI site is treated with MClaI prior to cutting.

34. The method of claim 32 wherein the rare sequence is cut with DpnI.

35. The method of claim 32 wherein the rare sequence is present in a tandem array.

36. The method of claim 1 wherein the genome to be mapped is a bacterial genome.

37. The method of claim 36 wherein integration is achieved by insertion of the cassette into a cell containing the genomic DNA.

38. The method of claim 36 wherein the insertion is achieved by vector transmission of the cassette DNA.

39. The method of claim 38 wherein the vector is a transposon.

40. The method of claim 1 wherein the genomic DNA is plant genomic DNA.

41. The method of claim 40 wherein integration is achieved by insertion of the cassette into a cell containing the genomic DNA.

42. The method of claim 41 wherein the insertion is achieved by vector transmission of the cassette DNA.

43. The method of claim 42 wherein the vector is a Ti plasmid.

44. The method of claim 1 wherein the genomic DNA is yeast DNA.

45. The method of claim 44 wherein integration is achieved by insertion of the cassette into a cell containing the genomic DNA.

46. The method of claim 45 wherein the insertion is achieved by vector transmission of the cassette DNA.

47. The method of claim 46 wherein the vector is a Ty element.

48. A DNA cassette comprising a restriction sequence rare in the vertebrate genome, flanked on at least one side by bacterial or viral DNA sequence.

49. The cassette of claim 48 wherein the vertebrate is a mammal.

50. The cassette of claim 49 wherein the flanking sequence is a viral sequence.

51. The cassette of claim 50 wherein the flanking sequence is a retroviral sequence.

52. The cassette of claim 51 wherein the flanking sequence is substantially homologous to a sequence found in the vector pZipNeo.

53. The cassette of any one of claims 48-52 wherein the rare restriction sequence is ClaI/ClaI.

54. The cassette of claim 48 wherein the rare restriction sequence is present in a tandem array.

55. The cassette of claim 53 wherein the rare restriction sequence is present in a tandem array.

56. The cassette of claim 53 wherein the ClaI/ClaI sequence has been methylated by MClaI.

57. The cassette of claim 55 wherein the ClaI/ClaI sequence has been methylated by MClaI.

58. A DNA cassette comprising a restriction sequence rare in an insect genome; being flanked on at least one side by a P element DNA sequence.

59. The cassette of claim 58 wherein the insect is Drosophila.

60. The cassette of claim 58 or 59 wherein the flanking sequence is substantially homologous to a sequence found in the vector pUChsneo.

61. The cassette of claim 58 or 59 wherein the rare restriction sequence is a ClaI/ClaI restriction site.

62. The cassette of claim 60 wherein the rare restriction sequence is present in a tandem array.

63. The cassette of claim 6_ wherein the rare restriction sequence is present in a tandem array.

64. The cassette of claim 62 wherein the ClaI/ClaI sequence has been methylated by MClaI.

65. The cassette of claim 63 wherein the ClaI/ClaI sequence has been methylated by MClaI.

66. A DNA cassette comprising a restriction sequence rare in the plant genome flanked on at least one side by a Ti plasmid DNA sequence.

67. The cassette of claim 66 wherein the rare restriction sequence is present in a tandem array.

68. A DNA cassette comprising a restriction sequence rare in the yeast genome flanked on at least one side by a Ty element DNA sequence.

69. The cassette of claim 68 wherein the rare restriction sequence is present in a tandem array.

70. A DNA cassette comprising a restriction sequence rare in the bacterial genome flanked on at least one side by a transposon DNA sequence.

71. The cassette of claim 70 wherein the rare restriction sequence is present in a tandem array.

72. A vector comprising the cassette of any one of claims 48-52.

73. A vector comprising the cassette of claim 53.

74. A vector comprising the cassette of claim 55.

75. A vector comprising the cassette of claim 56.

76. A vector comprising the cassette of claim 58 or 59.

77. A vector comprising the cassette of claim 60.

78. A vector comprising the cassette of claim 61.

79. A vector comprising the cassette of claim 62.

80. A vector comprising the cassette of claim 63.

81. A vector comprising the cassette of claim 65.

82. A vector comprising the cassette of claims 66 or 67.

83. A vector comprising the cassette of claims 68 or 69.

84. A vector comprising the cassette of claims 70 or 71.

85. A vertebrate cell having integrated into its genome the cassette of any one of claims 48-52.

86. The cell of claim 85 wherein the vertebrate is a mammal.

87. The cell of claim 86 wherein the mammal is a human.

88. A vertebrate cell having integrated into its genome the cassette of any one of claims 48-52.

89. A vertebrate cell having integrated into its genome the cassette of claim 53.

90. A vertebrate cell having integrated into its genome the cassette of claim 54.

91. A vertebrate cell having integrated into its genome the cassette of claim 55.

92. A vertebrate cell having integrated into its genome the cassette of claim 56.

93. A vertebrate cell having integrated into its genome the cassette of claim 57.

94. An insect cell having incorporated into its genome the cassette of claim 58 or 59.

95. An insect cell having incorporated into its genome the cassette of claim 60.

96. An insect cell having incorporated into its genome the cassette of claim 61.

97. An insect cell having incorporated into its genome the cassette of claim 62.

98. An insect cell having incorporated into its genome the cassette of claim 63.

99. An insect cell having incorporated into its genome the cassette of claim 64.

100. An insect cell having incorporated into its genome the cassette of claim 65.

101. A plant cell having integrated into its genome the cassette of claim 66 or 67.

102. A yeast cell having integrated into its genome the cassette of claim 68 or 69.

103. A bacterial cell having integrated into its genome the cassette of claim 70 or 71.

104. A continuous genomic map prepared according to the method of claims 1-11 or 15.

105. A continuous genomic map prepared according to the method of claim 12.

106. A continuous genomic map prepared according to the method of claim 14.

107. A continuous genomic map prepared according to the method of claim 20.

108. A continuous genomic map prepared according to the method of claim 27.

109. A continuous genomic map prepared according to the method of claim 29.

110. A continuous genomic map prepared according to the method of claim 37.

111. A continuous genomic map prepared according to the method of claim 41.

112. A continuous genomic map prepared according to the method 44.

INSERT CLA I/CLA I SEQUENCE ↓ INTO DEFECTIVE RETROVIRUS

UNIQUE DNA A	RESTRICT SITE	UNIQUE DNA B
5' ————————	ATCGATCGAT	————————
3' ————————	TAGCTAGCTA	————————

TRANSFECT INTO ↓ TRANS PACKAGING CELL LINE

ISOLATE INFECTIVE ↓ RETROVIRAL PARTICLES

INFECT CELL LINE ↓ OF INTEREST MONOTONICALLY

GROW UP CLONAL ↓ CELL ISOLATES

GENOMIC DNA	UNIQUE DNA A	RESTRICT SITE	UNIQUE DNA B	GENOMIC DNA
----------	————————	ATCGATCGAT	————————	----------
----------	————————	TAGCTAGCTA	————————	----------

ISOLATE ↓ DNA

METHYLATE DNA ↓ WITH M.CLA I

CLEAVE CLAI/CLAI SEQUENCE WITH ↓ METHYLATION SPECIFIC ENZYNE DPN I

PARTIALLY DIGEST DNA WITH ↓ A SECOND RESTRICTION ENZYNE

ELECTROPHORESE CLEAVED ↓ AND METHYLATED DNA

BLOT SEPARATED DNA TO SUPPORT ↓ MEMBRANE (NITROCELLULOSE, NYLON, ETC)

PROBE MEMBRANE WITH ↓ UNIQUE SEQUENCE A OR B

REPEAT PROCEDURE WITH (N) ↓ DIFFERENT CLONAL ISOLATES

ASSEMBLE MAP

FIG. 1

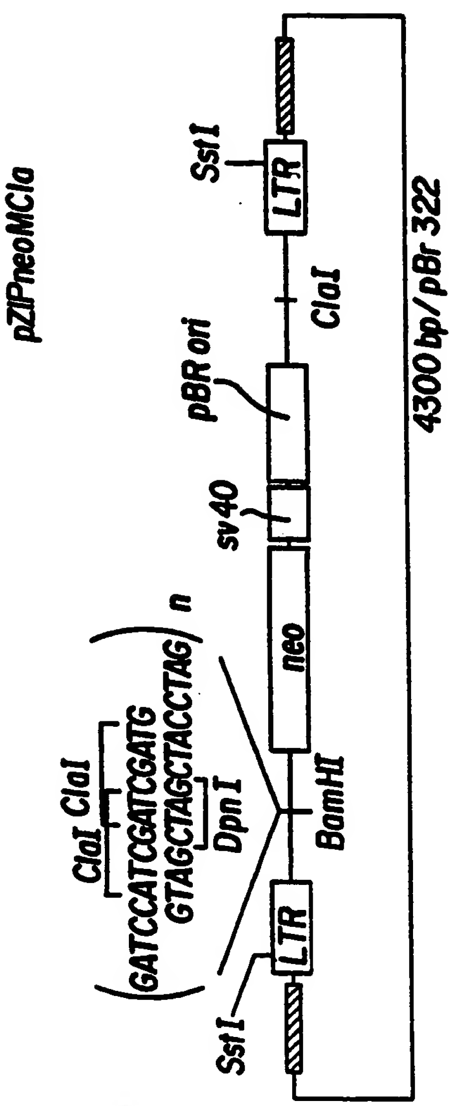

pZIPneoMCla
SstI
LTR
ClaI
pBR ori
sv40
neo
BamHI
ClaI ClaI
GATCCATCGATCGATG
GTAGCTAGCTACCTAG
DpnI
n
LTR
SstI
4300 bp/pBr 322

FIG. 2a

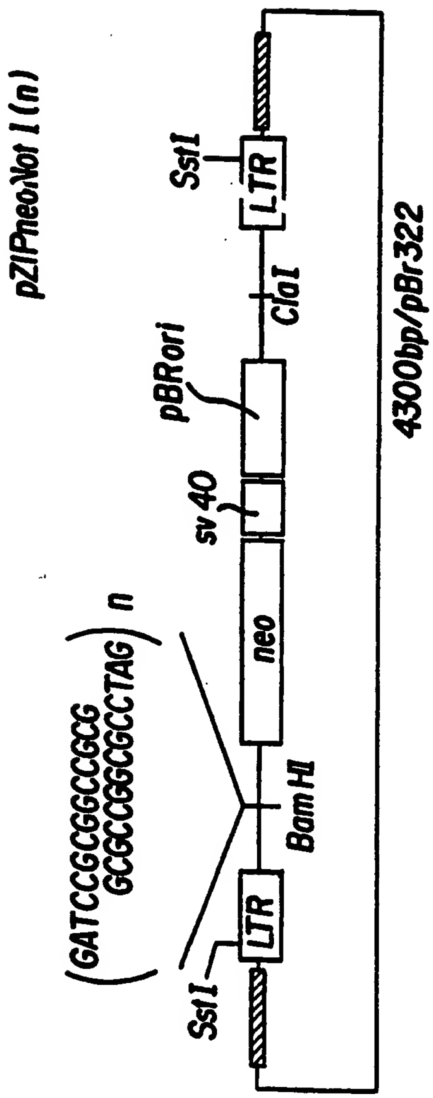

pZIPneo:Not I (n)
(GATCCGCGGCCGCG
GCGCCGGCGCCTAG) n
SstI
LTR
BamHI
neo
sv40
pBRori
ClaI
LTR
SstI
4300bp/pBr-322

FIG.2b

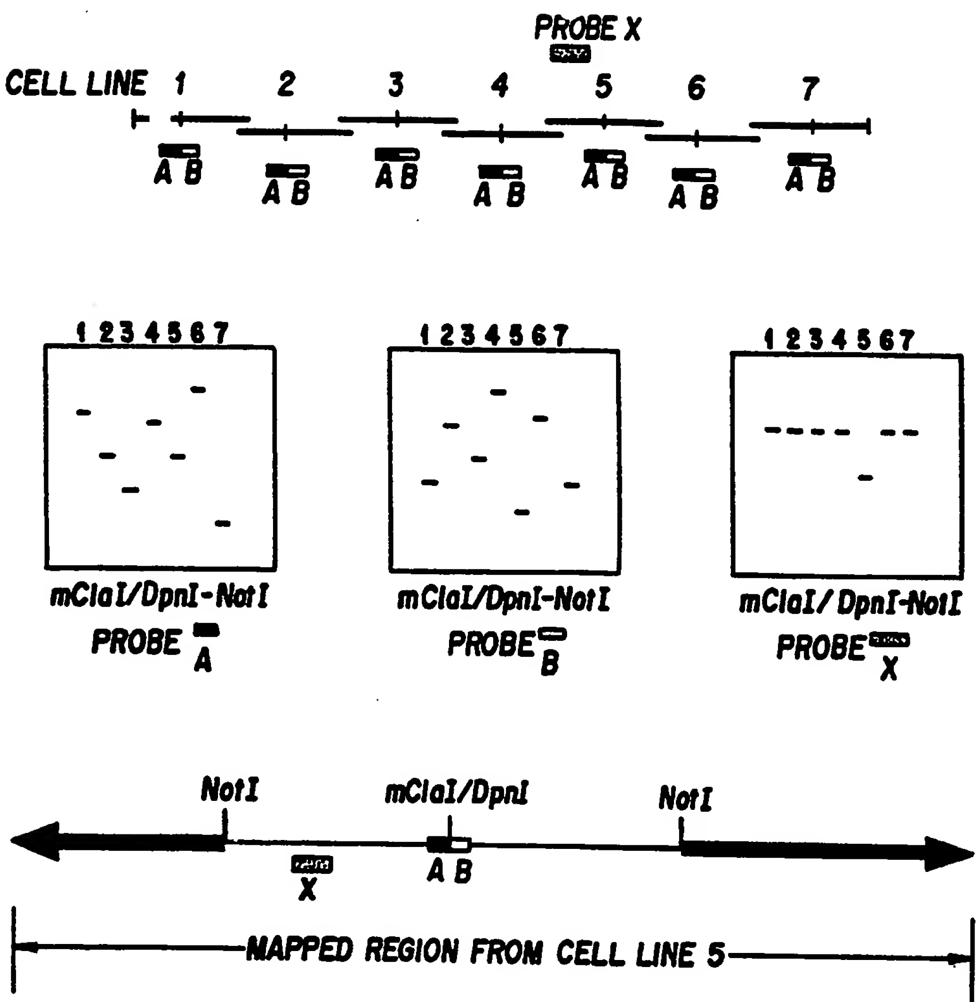

PROBE X
CELL LINE 1 2 3 4 5 6 7
A B
1 2 3 4 5 6 7
mClaI/DpnI-NotI
PROBE A
PROBE B
PROBE X
NotI
mClaI/DpnI
NotI
X
A B
MAPPED REGION FROM CELL LINE 5

FIG. 3

FIG. 4

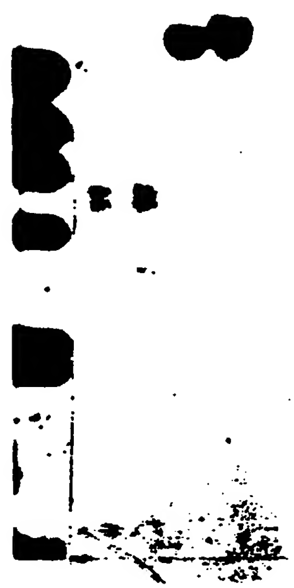
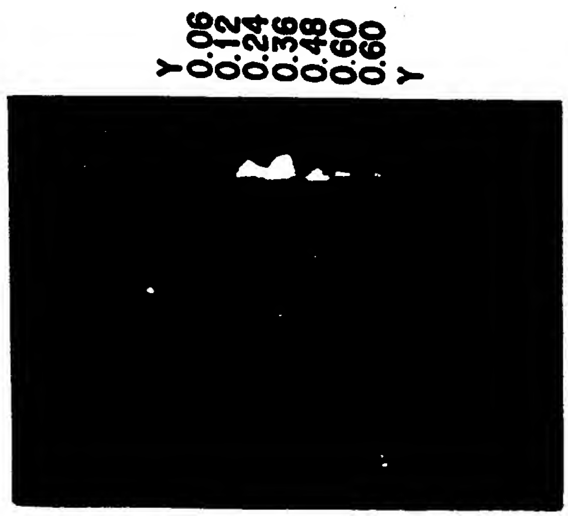


FIG. 5



INTERNATIONAL SEARCH REPORT

International Application No. PCT/US89/01983

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(4): C12Q 1/68; GOIN 33/53; C12N 5/00; C12N 7/00; C12P 19/34
US.Cl. 435/6; 536/27; 435/240.1; 435/235

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	435/6; 536/27; 435/240.1; 435/235

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

Computer Search APS;CAS; BIOSIS DATA BASES

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category [*]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X Y	Chemical Abstracts, Volume 109, No.21 issued November 1988 (Columbus, Ohio, USA), Garfinkel et al. "Transposon tagging using Ty elements in yeast", abstact No. 184405j;, Genetics (1988) 120(1) 95-108 see abstract	68,83,102 1-67,69-82 84-101,103-112
X Y	Chemical Abstracts, Volume 104, No.9 issued March 1986 (Columbus, Ohio, USA), Taylor et al. "Transcription of Agrobacterium rhizogenes A4 T-DNA, abstract No. 63016, Mol. Gen. Genet. (1985) 201(3) 546-53 see abstract	66,82,101 1-65,67-81,83-100, 102-112
A	American Journal of Human Genetics, Volume 32 1980 pages 314-331, (Chicago, Il. US) Botstein et al. "Construction of a Genetic Linkage Map Using Restriction Frangment Length Polymorphisms" see abstract	1-112

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
26 July 1989	12 SEP 1989
International Searching Authority	**Signature of Authorized Officer**
ISA/US	Scott A. Chambers

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	US,A, 4,173,452 (Cantor et al) 25 September 1984, See abstract	2
X Y	Gene, Volume 41 1986 pages 145-152, (amsterdan) Ubben et al. "Tn1721 derivative for transposon mutagenesis, restriction mapping and nucleotide sequence analysis" see abstract	70,84,103, 69,71-83 104-112
X Y	Cold Spring Harbor Symposia on Quantitative Biology, Volume 50 1985 page 439-445, Jaenisch et al. "Retroviruses and Insertional Mutagenesis" see Abstract	48-52, 72,85,86,88 1-47,53-71,73-84, 87,89-112
X Y	Cell, Volume 32 1983 pages 209-216, Jaenish et al. "Germline Integration of Moloney Murine Leukemia Virus at the Mov 13 Locus Leads to Recessive Lethal Mutation and Early Embryonic Death "see abstract	48-52,72,85 86,88 1-47,53-71,73-84,87,89-112
X Y	Science Volume 218 issued 1982 348-353 Rubin et al. "Genetic Transformation of Drosophila with Transposable Element Vectors" see abstract	48,49,58-60,72,77, 94,95 1-47,50-57 61-71,73-76